

学校编码: 10384
学号:

分类号
密级
UDC

厦门大学
硕士学位论文

P2P网络中的web服务发现机制研究

Study on Web Service Discovery Mechanism
in P2P Network

指导教师姓名: 教授

专 业 名 称:

论文提交日期: 2007 年 月

论文答辩时间: 2007 年 月

学位授予日期: 2007 年 月

答辩委员会主席:

评 阅 人:

2007 年 月

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1、保密（ ），在 年解密后适用本授权书。

2、不保密（ ）

（请在以上相应括号内打“√”）

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

摘要

随着 P2P 网络的迅猛发展,如何在庞大的 P2P 网络资源中快速准确地找到所需服务,已经成为 web 服务技术最关心的问题。然而,传统的服务发现机制存在着许多问题。尤其在搜索算法方面,主流的泛洪(Flooding)、随机漫步(Random walk)、Gnutella2 等算法都各有优劣。如何在保证稳定系统开销的基础上兼顾搜索速度和准确率,是 P2P 搜索算法最根本的问题。

本文在 P2P 网络模式、“小世界”现象和流言传播机制的基础上,提出了分布式电子商务服务发现模型 MatureNode。

首先,引入了“成熟节点”的概念,用节点“成熟度指标”使每个节点具有一定的“直觉”,而且用三种不同类型的“服务信息库”作为计算“成熟度指标”的依据。通过 OPNET 仿真实验与随机步搜索模型进行对比,证明该模型的效率有一定的提高。

其次,通过模拟流言的散布过程改进 MatureNode 模型,实现了广告转发机制。该机制主要包含两个关键参数:“转发概率”和“转发权值”。“转发概率”决定了广告将被转发的邻居节点数。与广告的广泛程度成反比,因此稀有的广告可以被迅速传播,而大众化的广告将被忽略;“转发权值”决定了广告将被转发到哪几个邻居节点。由邻居节点的成熟度指标 NMM、邻居节点与广告源节点之间的距离 D 两个因素决定。仿真实验从发现速度、系统开销、发现完整性三个方面验证了基于流言散布模型的服务广告转发机制的有效性:能够明显地提高服务发现的效率并保持稳定的系统开销,而且还能够在有效时间内发现更多的服务。

关键字: P2P; Web 服务发现; 流言传播.

Abstract

With the boom of P2P network, how to find out the needed services becomes the most concerned problem in web service technology field. However, there are a lot of problems in traditional service discovery mechanisms. All the popular discovery algorithms as flooding, random walk and Gnutella2 have both advantages and disadvantages. How to take care of discovery speed and accuracy rate and keep steady system cost at the same time, this is the root question of P2P searching algorithm.

On the base of P2P network model, “Small World Phenomena” and rumor spread mechanism, this paper puts forward the P2P e-commerce service discovery model MatureNode.

At first, this paper imports the conception of “Mature Node”. Using “Node Mature Metric (NMM)” to make every peer owns some kinder “instinct” and calculating NMM according to three different “service information bases”. Using OPNET simulation to compare with random walk algorithm, this algorithm has better efficiency.

The next step is simulating the process of rumor spreading to improve MatureNode model and built up a new ad forwarding mechanism. This ad forwarding mechanism contains two key parameters: “Forwarding Probability (P_f)” and “Forwarding Weight (W_f)”. P_f determines the number of neighbor nodes ads forwarded. It has inverted ratio with ad’s popularity, so rare ads will spread quickly and widely. W_f decides which group of neighbor nodes is chosen to forward ads. According to the results of simulation, on the base of steady system cost, MatureNode can obviously improve service discovery speed, furthermore it can discovery more services in provided TTL.

Keywords: P2P; Web Service Discovery; Rumor Spread Mechanism.

目录

第一章 绪论	1
1.1 研究背景	1
1.1.1 问题的提出	1
1.1.2 发展状况	1
1.2 论文主要创新点	2
1.3 论文组织结构	2
第二章 P2P 概述	3
2.1 P2P 介绍	3
2.2 P2P 网络的拓扑结构	5
2.2.1 中心化拓扑	6
2.2.2 全分布式非结构化拓扑	7
2.2.3 全分布式结构化拓扑	8
2.2.4 半分布式拓扑	10
2.3 P2P 搜索技术	12
2.3.1 结构化 P2P 网络 (DHT 网络) 的搜索技术	12
2.3.2 非结构化 P2P 网络的搜索技术	13
2.3.2.1 小世界模型 (Small World) 对 P2P 搜索技术的重大影响	13
2.3.2.2 非结构化 P2P 搜索算法	15
2.3.3 P2P 搜索技术研究的挑战	18
2.4 P2P 面临的问题与前景展望	19
2.4.1 面临的问题	19
2.4.2 前景展望	21
第三章 P2Ptalker 即时通讯软件	23
3.1 需求分析	23
3.2 系统分析	23
3.3 功能演示	23
第四章 基于成熟节点和流言机制的 P2P 电子商务服务发现模型	
MatureNode	27
4.1 流言传播机制介绍	27
4.1.1 流言机制的数学基础	27
4.1.2 流言机制的实验基础	28
4.1.3 基于流言机制的 P2P 发现过程	29
4.2 基于成熟节点的分布式电子商务服务发现模型	30
4.2.1 前提假设	30
4.2.2 成熟节点	31
4.2.3 服务信息库	31
4.2.4 成熟度指标	33
4.2.5 仿真实验及分析	34

4.3 基于流言模型的电子商务服务广告转发机制	38
4.3.1 服务发现模型的评价因素	39
4.3.2 流言散布模型	40
4.3.2.1 转发概率	40
4.3.2.2 转发权值	40
4.3.2.3 流言散布的广告转发算法	41
4.3.3 仿真实验及分析	41
第五章 总结	47
5.1 研究工作总结	47
5.2 进一步工作	47
附录 P2Ptalker 源程序	48
参考文献	59
致谢	61

Contents

1 Introduction	1
1.1 Study Background	1
1.1.1 Put Forward the Question	1
1.1.2 Development Status	1
1.2 New Ideas of This Paper	2
1.3 Structure of This Paper	2
2 P2P Overview	3
2.1 P2P Introduction	3
2.2 Topologies of P2P Network	5
2.2.1 Centralized Topology	6
2.2.2 Decentralized Unstructured Topology	7
2.2.3 Decentralized Structured Topology	8
2.2.4 Partially Decentralized Topology	10
2.3 P2P Search Technologies	12
2.3.1 Search Technologies in Structured P2P Network(DHT Network)	12
2.3.2 Search Technologies in Unstructured P2P Network	13
2.3.2.1 Important Influence of Small World Model to P2P Search Technologies	13
2.3.2.2 Unstructured P2P Search Algorithm	15
2.3.3 Challenge of P2P Search Technology Study	18
2.4 Problems P2P Faced and Foreground Prospect	19
2.4.1 Problems	19
2.4.2 Foreground Prospect	21
3 P2Ptalker Design	23
3.1 Requirement Analysis	23
3.2 System Analysis	23
3.3 Function Demo	23
4 Maturenode: P2P E-commerce Service Discovery Model Based on Mature-node and Rumour-spread Mechanism	27
4.1 Rumour-spread Mechanism Introduction	27
4.1.1 Mathematics Foundation of Rumour-spread Mechanism	27
4.1.2 Experiment Foundation of Rumour-spread Mechanism	28
4.1.3 P2P Discovery Process Based on Rumour-spread Mechanism	29
4.2 Decentralized E-commerce Service Discovery Model Based on Mature-node	30
4.2.1 Hypothesis	30
4.2.2 Mature Node	31
4.2.3 Service Information Base(SIB)	31
4.2.4 Node Mature Metric(NMM)	33

4.2.5 Simulation and Analysis	34
4.3 E-commerce Service Ad Forwarding Mechanism Based on Rumour-spread Model	38
4.3.1 Evaluation Factors of Service Discovery Model	39
4.3.2 Rumour-spread Model	40
4.3.2.1 Forwarding Probability	40
4.3.2.2 Forwarding Weight	40
4.3.2.3 Rumour-spread Ad Forwarding Algorithm	41
4.3.3 Simulation and Analysis	41
5 Summarize	47
5.1 Study Work Summarize	47
5.2 Next-step Work	47
Appendix	48
Reference	59
Appreciation	错误！未定义书签。

第一章 绪论

1.1 研究背景

1.1.1 问题的提出

Web 服务作为一种新兴的 Web 应用模式，是一个崭新的分布式计算模型，是 Web 上数据和信息集成的有效机制。从电子商务应用领域来看，复杂的应用连接和程序代码造成了电子商务应用的高维护代价和高更新代价，而 Web 服务正好能够解决这一问题，成为目前应用环境中最为合理的解决方案^[1, 2]。

传统互联网实现了计算机硬件的连通，Web 实现了网页的连通，Web 服务则实现了程序和程序之间的共享。基于 Web 的各种应用为客户提供了不同类型的 Web 服务，软件组件或应用程序都可以发布为服务，而这些服务可能具有不同的形式和不同的复杂程度。客户以某种方式在这些不同类型的 Web 服务中找到需要的服务以执行 Web 服务请求，就是所谓 Web 服务发现，Web 服务发现是 Web 服务系统架构中的一个重要部分。

要在大量的服务中，快速准确甚至自动化地发现服务并调用执行，良好的服务发现机制是实现目标的前提。P2P 是未来网络模式的大势所趋，如何在 P2P 网络中进行快速准确的服务发现，是很有前景的研究方向。本文提出了基于成熟节点和流言理论的分布式电子商务服务发现模型 MatureNode。

1.1.2 发展状况

目前，P2P 计算技术正不断应用到军事、商业、政府信息、通讯等领域。根据具体应用不同，可以把 P2P 分为大致以下这些类型：

- (1) 文件内容共享和下载，例如 Napster、Gnutella、eDonkey、eMule、Maze、BT 等；
- (2) 计算能力和存储共享，例如 <http://setiathome.ssl.berkeley.edu/>、Avaki、Popular Power 等；
- (3) 基于 P2P 技术的协同与服务共享平台，例如 JXTA、Magi、Groove 等；
- (4) 即时通讯工具，例如 ICQ、QQ、Yahoo Messenger、MSN Messenger 等；
- (5) P2P 通讯与信息共享，例如 Skype、Crowds、Onion Routing 等；
- (6) 基于 P2P 技术的网络电视：例如沸点、PPStream、PPLive、QQLive、SopCast

等。

1.2 论文主要创新点

(1)在 P2P 基本通信理论基础上,制作了 P2Ptalker 即时通讯软件。

(2)在 P2P 网络模式、“小世界”现象和流言传播机制的基础上,引入“成熟度指标”、“服务信息库”、“转发权值”、“转发概率”等关键概念,提出了分布式电子商务服务发现模型 MatureNode。

1.3 论文组织结构

本文的各章节结构和主要内容安排如下:

第一章 绪论:介绍了论文研究的背景、意义、国内外研究现状、论文主要创新点和结构安排。

第二章 P2P 概述:介绍 P2P 基本原理和四种拓扑结构,总结当前 P2P 各种搜索技术的成果、不足和面临的问题,并对 P2P 发展前景进行了展望。

第三章 P2Ptalker 即时通讯软件设计:介绍 P2Ptalker 设计中的需求分析、系统分析,并进行了功能演示。

第四章 基于成熟节点和流言机制的 P2P 电子商务发现模型 MatureNode:简要介绍流言传播机制。阐述 MatureNode 模型的关键概念、体系结构和搜索算法,通过仿真进行性能评价。

第五章 总结:研究工作总结及进一步工作介绍。

附录: P2Ptalker 源程序。

第二章 P2P 概述

2.1 P2P 介绍^[3]

千禧年伊始, Peer-to-Peer(对等计算, 简称 P2P)迅速成为计算机界关注的热门话题之一, 财富杂志更将 P2P 列为影响 Internet 未来的四项科技之一。

“Peer”在英语里有“对等者”和“伙伴”的意义。因此, 从字面上, P2P 可以理解为对等互联网。国内的媒体一般将 P2P 翻译成“点对点”或者“端对端”, 学术界则统一称为对等计算。P2P 可以定义为: 网络的参与者共享他们所拥有的一部分硬件资源(处理能力、存储能力、网络连接能力、打印机等), 这些共享资源通过网络提供服务 and 内容, 能被其它对等节点(Peer)直接访问而无需经过中间实体。在此网络中的参与者既是资源(服务和内容)提供者(Server), 又是资源获取者(Client)。

客观地说, 这种计算模式并不是什么新技术, 自从上个世纪 70 年代网络产生以来就存在了, 只不过当时的网络带宽和传播速度限制了这种计算模式的发展。90 年代末, 随着高速互联网的普及和个人计算机计算存储能力的提升, P2P 技术重新登上历史舞台并且带来了一场技术上的革命。许多基于 P2P 技术的杀手级应用应运而生, 给人们的生活带来了极大的便利。

从计算模式上来说, P2P 打破了传统的 Client/Server(C/S)模式, 在网络中的每个节点的地位都是对等的。每个节点既充当服务器, 为其他节点提供服务, 同时也享用其他节点提供的服务。P2P 与 C/S 模式的对比如下图所示:

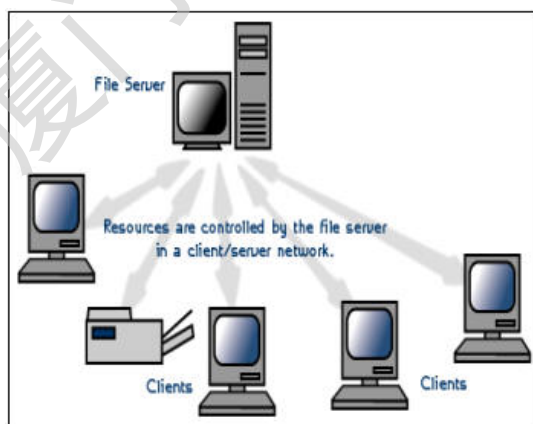


图 2-1 Client/Server 模式

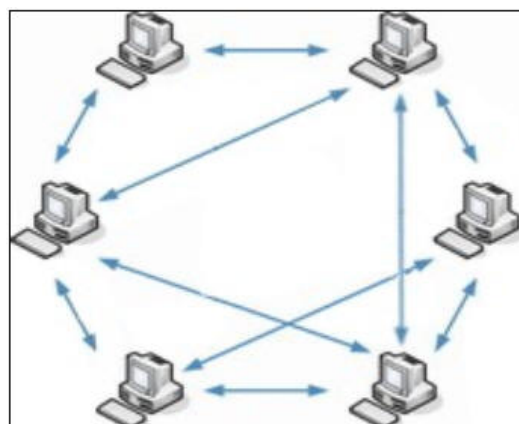


图 2-2 Peer-to-Peer 模式

P2P 技术的特点体现在以下几个方面：

(1) **非中心化**：网络中的资源和服务分散在所有节点上，信息的传输和服务的实现都直接在节点之间进行，可以无需中间环节和服务器的介入，避免了可能的瓶颈。P2P 的非中心化基本特点，带来了其在可扩展性、健壮性等方面的优势。

(2) **可扩展性**：在 P2P 网络中，随着用户的加入，不仅服务的需求增加了，系统整体的资源和服务能力也在同步扩充，始终能比较容易地满足用户的需要。理论上其可扩展性几乎可以认为是无限的。例如：在传统的通过 FTP 的文件下载方式中，当下载用户增加之后，下载速度会变得越来越慢，然而 P2P 网络正好相反，加入的用户越多，P2P 网络中提供的资源就越多，下载的速度反而越快。

(3) **健壮性**：P2P 架构天生具有耐攻击、高容错的优点。由于服务是分散在各个节点之间进行的，部分节点或网络遭到破坏对其它部分的影响很小。P2P 网络一般在部分节点失效时能够自动调整整体拓扑，保持其它节点的连通性。P2P 网络通常都是以自组织的方式建立起来的，并允许节点自由地加入和离开。

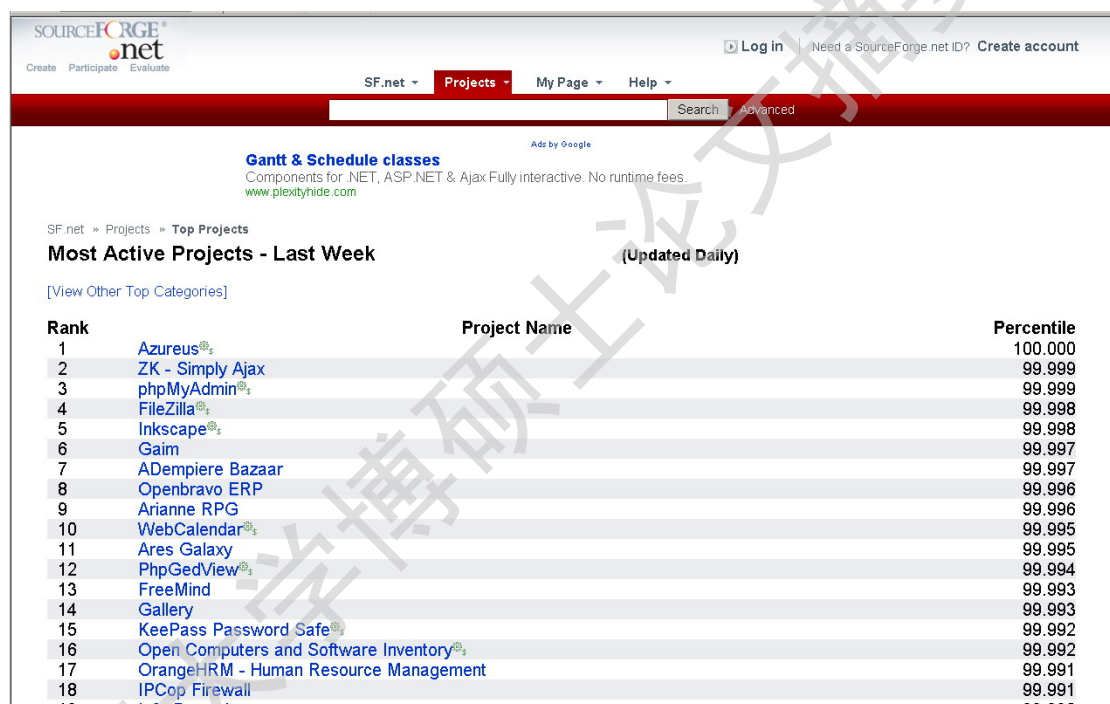
(4) **高性价比**：性能优势是 P2P 被广泛关注的一个重要原因。随着硬件技术的发展，个人计算机的计算和存储能力以及网络带宽等性能依照摩尔定理高速增长。采用 P2P 架构可以有效地利用互联网中散布的大量普通节点，将计算任务或存储资料分布到所有节点上。利用其中闲置的计算能力和存储空间达到高性能计算和海量存储的目的。目前，P2P 在这方面的应用多在学术研究方面，一旦技术成熟能够在工业领域推广，则可以为许多企业节省购买大型服务器的成本。

(5) **隐私保护**：在 P2P 网络中，由于信息的传输分散在各节点之间进行而无需经过某个集中环节，用户的隐私信息被窃听和泄漏的可能性大大缩小。此外，目前解决 Internet 隐私问题主要采用中继转发的技术方法，从而将通信的参与者隐藏在众多的网络实体之中。在传统的一些匿名通信系统中，实现这一机制依赖于某些中继服务器节点；而在 P2P 中，所有参与者都可以提供中继转发的功能，因而大大提高了匿名通讯的灵活性和可靠性，能够为用户提供更好的隐私保护。

(6) **负载均衡**：P2P 网络环境下由于每个节点既是服务器又是客户机，减少了传统 C/S 结构中对服务器计算能力、存储能力的要求，同时资源分布在多个节点，所以可以更好地实现整个网络的负载均衡。

与传统的分布式系统相比，P2P 技术具有无可比拟的优势，同时，P2P 技术

具有广阔的应用前景。目前, Internet 上各种 P2P 应用软件层出不穷, 用户数量急剧增加。据统计, 自 2001 年以来, 大量 P2P 软件的用户使用数量从几十万、几百万到上千万急剧增加, 给 Internet 带宽带来巨大冲击。如图 2-3 所示, 在全球最大的开源网站 Sourceforge^[4] 下载排名中, 前十五名中有十一个项目是基于 P2P 技术。其中 Azureus 的下载量超过 1 亿次。这个网站上的工程项目下载量, 往往反映当今软件技术的前沿热点。令人瞩目的是, 微软公司在新一代操作系统 Windows Vista 中也加入了 P2P 技术用来加强协作和应用程序之间的通讯^[5]。



Rank	Project Name	Percentile
1	Azureus	100.000
2	ZK - Simply Ajax	99.999
3	phpMyAdmin	99.999
4	FileZilla	99.998
5	Inkscape	99.998
6	Gaim	99.997
7	ADempiere Bazaar	99.997
8	Openbravo ERP	99.996
9	Arianne RPG	99.996
10	WebCalendar	99.995
11	Ares Galaxy	99.995
12	PhpGedView	99.994
13	FreeMind	99.993
14	Gallery	99.993
15	KeePass Password Safe	99.992
16	Open Computers and Software Inventory	99.992
17	OrangeHRM - Human Resource Management	99.991
18	IPCop Firewall	99.991

图 2-3 Sourceforge 工程下载排名 (数据截止到 2007 年 3 月 23 日)

2.2 P2P 网络的拓扑结构^[3]

拓扑结构是指分布式系统中各个计算单元之间的物理或逻辑的互联关系, 节点之间的拓扑结构一直是确定系统类型的重要依据。目前互连网络中广泛使用集中式、层次式等拓扑结构。Internet 本身是世界上最大的非集中式的互连网络, 但是九十年代所建立的一些网络应用系统却是完全的集中式的系统, 许多 Web 应用都是运行在集中式的服务器系统上。集中式拓扑结构系统目前面临着过量存储负载、DOS (Denial of Service, 拒绝服务) 攻击、网络带宽限制等一些难以解

决的问题。Peer-to-Peer（简称 P2P）系统主要采用非集中式的拓扑结构，一般来说不存在上述这些难题。根据结构关系可以将 P2P 系统细分为四种拓扑形式：

- (1) 中心化拓扑（Centralized Topology）；
- (2) 全分布式非结构化拓扑（Decentralized Unstructured Topology）；
- (3) 全分布式结构化拓扑（Decentralized Structured Topology，也称作 DHT 网络）；
- (4) 半分布式拓扑（Partially Decentralized Topology）。

2.2.1 中心化拓扑

中心化拓扑最大的优点是维护简单，资源发现效率高。由于资源的发现依赖中心化的目录系统，发现算法灵活高效并能够实现复杂查询。最大的问题与传统 C/S 结构类似，容易造成单点故障、“热点”现象、版权纠纷等问题。这是第一代 P2P 网络采用的结构模式，经典案例就是著名的 MP3 共享软件 Napster^[6]。

Napster 是最早出现的 P2P 系统之一，并在短期内迅速成长起来。它实质上并非纯粹的 P2P 系统，而是通过一个中央索引服务器保存所有 Napster 用户上传的音乐文件索引和存放位置的信息。它的工作原理如图 2-4 所示。当某个用户需要某个音乐文件时，首先连接到 Napster 中央索引服务器，在服务器上进行搜索，服务器返回存有该文件的用户信息，再由请求者直接连到文件的所有者传输文件。Napster 首先实现了文件查询与文件传输的分离，有效地节省了中央服务器的带宽消耗，减少了系统的文件传输延时。

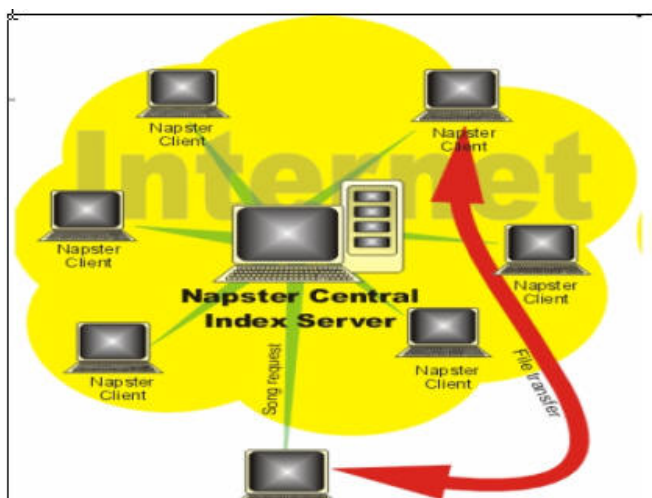


图 2-4 Napster 的拓扑结构

然而, 这种对等网络模型存在以下这些问题:

(1) 中央索引服务器的瘫痪容易导致整个网络的崩溃, 因此可靠性和安全性较低。

(2) 随着网络规模的扩大, 对中央索引服务器进行维护和更新的费用将急剧增加, 所需成本较高。

(3) 中央索引服务器的存在常引起版权问题上的纠纷, 服务提供商容易被追究法律责任。

综合上述优缺点, 对小型网络而言, 中心化拓扑模型在管理和控制方面占一定优势。但鉴于其存在的上述缺陷, 该模型并不适合大型网络应用。

2.2.2 全分布式非结构化拓扑

全分布式非结构化拓扑的 P2P 网络是在重叠网络 (Overlay Network) 中采用了随机图的组织方式, 节点度数服从 Power-law 规律 (幂次法则)^[7], 能够较快发现目的节点, 面对网络的动态变化体现了较好的容错能力, 因此具有较好的可用性。同时可以支持复杂查询, 如带有规则表达式的多关键词查询、模糊查询等, 采用这种拓扑结构最典型的案例便是 Gnutella (音译: 纽特拉)。准确地说, Gnutella 不是特指某一款软件, 而是指遵守 Gnutella 协议^[8]的网络以及客户端软件的统称。目前基于 Gnutella 网络的客户端软件非常多, 著名的有 Shareaza、LimeWire 和 BearShare 等。

Gnutella 和 Napster 最大的区别在于 Gnutella 是更加纯粹的 P2P 系统, 因为它没有中央索引服务器, 每台机器在 Gnutella 网络中是真正的对等关系, 既是客户机同时又是服务器, 所以被称为对等机 (Servent, Server+Client 的组合)。在文件检索方面, 它与 Napster 也不相同。在 Gnutella 网络的发展初期, 它主要采用基于完全随机图的 Flooding (泛洪) 搜索算法。图 2-5 显示了 Flooding 的工作流程: 当一台计算机要下载一个文件, 它首先以文件名或者关键字生成一个查询, 并把这个查询发送给与它相连的所有计算机, 这些计算机如果存在这个文件, 则与查询的机器建立连接; 如果不存在这个文件, 则继续在自己相邻的计算机之间转发这个查询, 直到找到文件为止。为了控制搜索消息不至于永远这样传递下去, 一般通过 TTL (Time to Live, 生存期) 的减值来控制查询的深度。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库